

Weekly Report

Junhua Lu

May 31, 2015

1. Finished graduation thesis defense.
2. We failed to extract the data but fortunately the oracle database recovers and we will process the data on Monday.
3. In logistic regression, missing value is not a easy problem to deal with, we may have to simulate reasonable values to compensate. This problem should be taken into consideration: either extract only persons with complete data or doing pre-process before regression.
4. In matlab, GLM function supports logistic regression so it's easy to process data. Decision tree is also not difficult to implement. However, there also exists problems in missing values. And there is also a method named "Random Forest", which was mentioned in someone's master thesis about Netease Data. It's a generalizing of decision tree.
5. In duration modal, there is another modal named Cox regression, it is also widely used. Different from Prof. He's paper using log-likelihood to estimate parameters, it uses simply regression to fit.
6. Feng YF says He's method is difficult in estimating since there are too much parameters. He concentrates on numerical algorithms to get the optimal solutions. KE JM are finding other statistical ways like wavelet analysis.
7. Next Wednesday I will go on a graduation trip.